

Here is the information on the problem with the historical data load problem.

Background:

We have discovered three problems with FY94 - to early FY99 (Dec 1998) APC data. When these data were loaded from the mainframe to the Informix database, problems occurred as shown in the next section. The problems appear to be isolated to historical records that were loaded with a specific program. This program is not the main program we use to process the PCC data received in the current PCC exports. We have corrected the problem in the program if we need to use this program in the future to process historical records.

Problems:

1. For most of the social security numbers (SSNs), instead of loading all 9 digits, the program loaded the SSN by truncating the first digit, loading the 2-9 digits, and then loading the first digit of the secondary provider code, as shown below:

- Where secondary provider code was null, it loaded the last digit of the SSN as a space.

Example: SSN on exported record: 565174321
 Secondary Prov Code: (blank)
 Loaded SSN: 65174321

- Where secondary provider code began with "0" through "9," it loaded the first digit of the secondary provider code as the last digit of the SSN.

Example: SSN on exported record: 565174557
 Secondary Prov Code: 00
 Loaded SSN: 651745570

2. For some records (depends on the value immediately before the SSN value) a problem with the conversion from EBCDIC to ASCII occurred, which caused only the first 4 digits of the SSN to be loaded.

3. For any fields that occurred after the SSN, the values are corrupted. We have confirmed that none of these fields are used to determine if the visit is workload reportable; however, the following fields most likely have inaccurate values:

- ICD9 codes. Should be minimal impact for FY94 - FY97 data since for the majority of visits we did not receive ICD9 codes. We've been told that in approximately May 1998, IHS started actively using/sending ICD9 codes to NPIRS; thus, there could be more of an impact on FY98 data and early FY99 data that was loaded using this program.

- Secondary provider codes.
- Surgical procedure ICD9 codes.

Impact Scenarios:

In all three problems, it appears we have two scenarios for these records: (1) either we received the record again in another export and the record was processed correctly and has the correct SSN or (2) we never received the record again.

Scenario 1, Received Record Again in Another Export

If we were to receive the same record in another export, which processes the record correctly, one of two circumstances would occur, as shown below. One example of when we would receive the record again is if the Area re-exported the visit in PCC, which has been done for FY98 and FY99 data for some facilities.

A. For records with chart number equal to "999999," the program would have to unduplicate the record by patient's SSN. Because the SSN on the record that was processed correctly is different from the SSN on the record that was processed incorrectly, the program treated the record as a new record and inserted it. So, now we have two records for the same visit, and if the visits met the workload reportable criteria, which if one did the other did, we have two workload reportable visits for the same visit. We believe there is no impact on the user population report because in this scenario, one of these visits is linkable to the patient and that is the one with the correct SSN. However, there is a problem with duplicate workload counts when the SSN was loaded as described in the Problems section #1 and #2 above.

The following information is what we have available for SSNs that were loaded by truncating the first digit, loading the 2-9, and loading the first digit of the secondary provider code as the 9th digit of the SSN. We do not have counts on the problem where only the first 4 digits of the SSN was loaded.

- FY98 total number of workload reportable duplicate records is 102,979, with Alaska Area being hardest hit. They had 65,426 records, or approximately 7.4% of their total APC workload for FY98. That means if we were to delete these duplicate visits, their FY98 APC workload would decrease by 7.4%. The problem does appear to exist for all Areas, but none of the Areas were affected as much as Alaska.
- FY99 total number of workload reportable duplicate records is 5,849, and once again Alaska Area was hardest hit. They had 3,594 of the 5,849 records. The problem does appear to exist for all Areas, but none of the Areas were affected as much as Alaska.
- FY97 had 2,269 total workload reportable dups; however, FY97 workload is officially verified and the only thing we would need to do would be to delete the duplicates.

B. For records with chart numbers not equal to "999999," the program would overwrite the existing, bad record. Thus the data in the SSN, ICD9 codes, secondary provider codes, and surgical procedure code fields would be corrected.

Scenario 2, Never Received Record Again

A. For all records, the data in the SSN, ICD9 codes, secondary provider codes, and surgical procedure code fields are not reliable. The only way to correct them is to reload. We anticipate the majority of the records fall into this scenario.

B. In this scenario, if the patient's chart number is "999999," when we try to link the visit to a patient, we are unable to because of the bad SSN. Therefore, the visit is not considered when determining if that patient is active for user population purposes. If these visits were reloaded correctly, they could contribute to slightly higher active user population counts. We ran a query to assess the extent of this problem and the results were as follows.

FY97 - 99 Workload Reportable Visits, Chart Numbers = "999999," Currently Not Linked to Patients:

Aberdeen	3,607
Alaska	33,440
Albuquerque	682
Bemidji	4,909
Billings	6,231
California	530
Nashville	2,788
Navajo	2,565
Oklahoma	37
Portland	759
Phoenix	2,434
<u>Tucson</u>	<u>338</u>
TOTAL:	58,320

Alaska Area would have the most opportunity to benefit from reloading this data as they had the most visits that are not linked to patients that have the opportunity to be linked if these visits were reloaded.

C. In this scenario, if the patient's chart number is not equal to "999999" but no match is made to the patient's chart number, the program then tries to link by SSN. We do know this has happened and it results in the visit not being linked to the patient; thus, the visit is not used for determining if the patient is an active user for user population reporting purposes. (NOTE: We are working on correcting the problem of patients that don't have charts in the NPIRS database.) We ran a query to assess the extent of this problem and the following are the results.

**FY97 - 99 Workload Reportable Visits, Chart
Numbers Not Equal to "999999," Currently Not Linked to Patients:**

All Areas 175,340

Phoenix Area had approximately 97% of these visits, so they are most likely to see any increases in active user population counts.

What to do to Fix the Problems:

At this time it looks like we can identify all of these records by the values in the Row_Creator and Source fields. Our recommendations for correcting this problem are shown below:

1. Focus on correcting the FY97 through FY99 workload data first, since that is the data used for FY99 User Population. Agree to fix the other data as soon as possible (steps #2 - 8 below would be repeated when correcting this data).
2. Run FY97 through FY99 APC 1A and FY99 User Population reports first so we know what our counts look like before we correct the data.
3. Get total record counts on the APC table.
4. Identify and copy the "bad" data to a separate work table so that it may be used for reference or reload if necessary.
5. Delete the "bad" data from the APC table.
6. Reload the FY96 through FY98 historical data from the mainframe master files. FY96 is included because it contains early FY97 data that is not included in the FY97 mainframe master files.
7. Get total record counts on the APC table.
8. Re-run FY97 through FY99 APC 1A and FY99 User Population reports and compare with "before" versions to see if outcome is as expected. If it isn't, research the differences.

Other Information:

We checked the separate programs used to load the historical Inpatient, CHS Inpatient, and CHS Outpatient tables, and the problem did not occur in them. We will check the program used to load historical Dental data next week.

Other Issues:

1. In the past, we did NOT maintain the original workload reportable flag of the record. This means that if we were to run an FY97 APC 1A report today, it would not match the counts it did back then. There are several factors contributing to this problem:
 - A. We have reset the workload reportable flags using current workload reporting logic, which could be different than it was in the past.
 - B. We basically have no cutoff logic implemented. Therefore, records that were not received and originally counted in FY97, for example, would now be counted.

We now have an opportunity to fix this. We have the original workload flag in the mainframe master files. We are recommending that we add and populate a field to the APC table, such as `original_workload_flag`. Then for data that has been officially verified, currently which is FY94 - FY97 for APC, when we run a workload report, we count only those visits that have a value in that field equal to "Y." In the example in item B above, all other FY97 APC visits in our database would not be counted.

2. This issue is somewhat related to issue #1 above. For visits associated to a fiscal year that has already been officially verified, our programs currently will allow the visit to be overwritten with new information if the visit were to be re-exported. This could potentially change any value in the record if the new record had different information. For example, if the old record had clinic code 20 and the new record has clinic code 30, the record would reflect clinic code 30. Our recommendation is that any record for which the fiscal year has been officially approved, the record should not be altered at all. The questions we are looking to resolve now are: What do we do with the new record? Do we store it in another table or do we bypass the record altogether?